Linked Conservation Data
Sharing Conservation Vocabularies

Kristen St John
Athanasios Velios

NKOS 2020
9 September 2020
Heritage (not biodiversity)
Documentation a critical task
Overall aim:
• Make conservation documentation widely accessible

Three project strands:
• Outreach and Education
• Terminology
• Modelling

Use Linked Open Data!
Linked Conservation Data - Thesaurus and Glossary Usage Questionnaire

Email address *
Valid email address
This form is collecting email addresses. Change settings

Name *
Short-answer text

Organization
Short-answer text

What is your current professional role? *
- Conservator
- Conservation Scientist
- Thesaurus editor
- Data manager
- Other...

What types of documentation do you generate? (check all that apply) *
- Conservation Treatment Documentation
- Other...
<table>
<thead>
<tr>
<th>E</th>
<th>F</th>
<th>G</th>
<th>H</th>
<th>I</th>
<th>J</th>
<th>K</th>
<th>L</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>What is your current professional role? What types of documents do you work on? Is your work primarily...</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conservator</td>
<td>Conservation Treatment</td>
<td>No</td>
<td>Published thesaurus - if yes, please proceed to the following question to indicate which ones.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assistant professor and/or Scientific Data Collection</td>
<td>Yes</td>
<td>Published thesaurus - if yes, please proceed to Heritage Conservation Terminology Definition of terms from various sources.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conservator</td>
<td>Conservation Treatment</td>
<td>No</td>
<td>No controlled vocabulary, choices guided by experience and knowledge.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conservator</td>
<td>Conservation Treatment</td>
<td>No</td>
<td>No controlled vocabulary, choices guided by experience and knowledge.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conservator, Lecturer</td>
<td>Conservation Treatment</td>
<td>Yes</td>
<td>No controlled vocabulary, choices guided by experience and knowledge.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conservator</td>
<td>Conservation Treatment</td>
<td>Yes</td>
<td>Published thesaurus - if yes, please proceed to Art &amp; Architecture Thesaurus. Getty Research Institute.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conservator</td>
<td>Conservation Treatment</td>
<td>No</td>
<td>No controlled vocabulary, choices guided by experience and knowledge.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conservator</td>
<td>Conservation Treatment</td>
<td>No</td>
<td>No controlled vocabulary, choices guided by experience and knowledge.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conservator</td>
<td>Conservation Treatment</td>
<td>Yes</td>
<td>In house glossary or thesaurus.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conservator</td>
<td>Conservation treatment</td>
<td>No</td>
<td>No controlled vocabulary, choices guided by experience and knowledge.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conservator</td>
<td>Conservation Treatment</td>
<td>No</td>
<td>Published thesaurus - if yes, please proceed to European or British Standards such as: EN 15986, EN 16085, e Glossary of Terms: American Institute for Conservation. Book and Paper Group.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Conservation controlled vocabularies

A library, media, and archival preservation glossary
In category: paper, books

ABC of bookbinding
In category: books

AIC BPG Glossary of Terms
Glossary of the paper conservation group of the American Institute for Conservation
In category: books, paper

AIC Lexicon Terms
In category: general

Art & Architecture Thesaurus (AAT)
The AAT is a thesaurus containing generic terms, dates, relationships, sources, and notes for work types, roles, materials, styles, cultures, techniques, and other concepts related to art, architecture, and other cultural heritage (e.g., amphora, oil paint, oilerverf, acetoysis, sintering, orthographic drawings, Olmeca, Rinascimento, Buddhism, watercolors, asa-no-ha-toji, sralais).
In category: general

Binding Terms. Rare Books and Manuscripts Section: Controlled Vocabularies
Terminology workshop (pre COVID-19!)
Workshop outcomes

• Wide range of vocabularies: from LOD SKOS thesauri to unsorted wordlists without IDs
• Good coverage: materials, object components
• Poor coverage: conditions, treatment techniques
• Getty Arts & Architecture Thesaurus
  – hub for aligning vocabularies
  – best candidate to expand for coverage
• Backbone Thesaurus
  – hub for CIDOC-CRM modelling
textblocks

Preferred label:
textblocks
tekstblokk

The word textblock is used to describe all the leaves in a book on which the text is written or printed. Where there is more than one text in a single binding, as in the case of composite volumes or Sammelbände, all the different texts are included in a single textblock. A textblock does not include endleaves or other leaves added by a binder, such as inserted text separators or interleaving, even though these may now carry additional written material. Endleaves and all other leaves added by the binder are, however, included in a bookblock. In certain exceptional circumstances, such as dos-à-dos bindings, a single binding may contain two or more bookblocks, each of which can in theory be a composite volume. Stationery bindings will often be made with a textblock which consists of blank gatherings yet to be written in, in which the outermost leaf of the outermost gathering at each end will be used as endleaves in the form of pastedowns (i.e. integral endleaves), and in these cases the textblock and the bookblock are the same thing. If the outermost gatherings of such a book are made in a different format from the rest of the gatherings (e.g. four leaves as opposed to eight leaves, or outside hooks instead of bifolia), or made from a different, possibly coloured, paper, these can be described as endleaves, and the other leaves as the textblock, together making the bookblock.

Hierarchy:
<table>
<thead>
<tr>
<th>Hungarian</th>
<th>German</th>
<th>English</th>
<th>French</th>
<th>Italian</th>
</tr>
</thead>
<tbody>
<tr>
<td>acélmeteszés</td>
<td>Stahlstich</td>
<td>steel(-plate) engraving</td>
<td>gravure sur acier</td>
<td>incisione in acciaio</td>
</tr>
<tr>
<td>acélmetzet</td>
<td>Stahlstich</td>
<td>steel(-plate) engraving</td>
<td>gravure sur acier</td>
<td>incisione in acciaio</td>
</tr>
<tr>
<td>achátkő</td>
<td>Achatstein</td>
<td>agate burnisher</td>
<td>bruniross d’agate</td>
<td>pietra agata</td>
</tr>
<tr>
<td>acquaforte l. rézmetszet</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>akvarell</td>
<td>Aquarell, Gouache, Wasserfarben</td>
<td>aquarelle, gouache painting</td>
<td>couleur à l’eau, gouache</td>
<td>acquerello, colore a guazzo, guazzo</td>
</tr>
<tr>
<td>álbumdák</td>
<td>falsche Bündle</td>
<td>false bands</td>
<td>faux nerfs</td>
<td>falsa nervatura, nervi falsi, nervi finti</td>
</tr>
<tr>
<td>álcázott festett metszés</td>
<td>verschobener Schnitt</td>
<td>concealed for-edge painting</td>
<td>-</td>
<td>taglio mascherato</td>
</tr>
<tr>
<td>állag</td>
<td>Konsistenz</td>
<td>consistency</td>
<td>consistence</td>
<td>consistenza, sostanza</td>
</tr>
<tr>
<td>állítható körző l. mérőkörző</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>álló prés</td>
<td>Stockpresse</td>
<td>standing press</td>
<td>pressé à percussion (de reliure)</td>
<td>pressa a percussione</td>
</tr>
<tr>
<td>állomány (könyvtári)</td>
<td>Bestand, Buchbestand, Büchersammlung</td>
<td>book collection, bookstock, holding stock</td>
<td>fond (de livres) stock</td>
<td>collezione, fondo di libri, materiale librario</td>
</tr>
<tr>
<td>állományvédelem</td>
<td>Bestandeserhaltung</td>
<td>preservation</td>
<td>préservation</td>
<td>preservazione</td>
</tr>
<tr>
<td>ammónia</td>
<td>Ammoniak</td>
<td>anhydrous-ammonion</td>
<td>ammoniac</td>
<td>ammoniaca</td>
</tr>
<tr>
<td>ammónium-hidroxid</td>
<td>Ammoniakhydrat, Ammoniumhydroxid, Salmiakgeist</td>
<td>ammonia hydrate, ammonium hydroxide</td>
<td>hydroxyde d’ammonium</td>
<td>hidrossido di ammoniaca, hidrossido di salmico</td>
</tr>
<tr>
<td>ammónium-karbónát</td>
<td>Hirschhornsalz</td>
<td>carbonate of ammonia</td>
<td>carbonate d’ammonium, sel de corne de cerf</td>
<td>carbonato d’ammonio</td>
</tr>
<tr>
<td>anilinfesték, -szinezék</td>
<td>Anilinfarbe</td>
<td>aniline-colour, coal-tar dyestuff</td>
<td>couleur d’aniline</td>
<td>colore di anilina</td>
</tr>
</tbody>
</table>
Alignment

• Encoding vocabularies
  – to make them machine readable
How to publish as SKOS
For vocabulary maintainers

1. Encoding is when the concepts and labels from a vocabulary published as text (e.g. in a pdf format) are isolated manually and placed in some sort of data structure, e.g. a comma separated value (CSV) file (typically in a spreadsheet) or an XML file using markup.

2. Identifiers are strings of characters which point to a concept (or label). Encoded data often have local identifiers (unique in a local context), but these may not be suitable for Linked Data which require unique identifiers in a global context (also known as URIs). E.g. the URI for the concept of paper in LoB is: http://w3id.org/lob/concept1481

3. For a large vocabulary many hundreds or thousands of URIs will need to be published and maintained long term (e.g. in LoB the URI for the concept paper is this: http://w3id.org/lob/concept1481 currently pointing to https://www.lisatus.org.uk/lob/concept1481). A system to manage these is necessary. URIs may be needed for:
   - concepts
   - labels
   - datasets
   - vocabularies

   URIs may need to reflect versions. Consider keeping the same URI for a concept in different versions (the essence of a concept should not change), but change the dataset URI based on the version.

   The w3id consortium offers a service for URI redirection.

   Guidelines for design of URIs can be found online, e.g. here, here and here.

4. Mapping means matching the fields of the encoded CSV or XML file to SKOS RDF (as triples). Some of the tools mentioned in note 6 are able to import and convert to SKOS. Manual mapping can be done using tools such as JML, Kerko and STILETO.

5. Sharing means to make the resulting SKOS file available online (e.g. in a repository as RDF, turtle or RDF.XML) and/or to make the data available at a SPARQL endpoint for querying. LCD is discussing the adoption of a repository for SKOSified vocabularies see (??).

6. This step is optional: it is useful for humans to be able to look up terms in the vocabulary in addition to sharing as data. Tools to build such a website include BioPortal, Skosmos, Wordnet and TemaTrav. LCD is discussing the development of a conservation portal for SKOSified vocabularies see (??).

START

Is vocab published in a machine-friendly format?

- yes
  - Encode concepts with (at least) scope note and preferred label
  - Are there unique local and global identifiers?
    - yes
      - Map encoded file to SKOS
      - Produce local IDs and URIs and add to encoded file
      - Decide on local ID and URI structure
    - no
      - Share as downloadable files and/or at a SPARQL endpoint.

- no
  - Build website to serve vocab in a human friendly way
How to publish as SKOS
For vocabulary maintainers

START

Is vocab published in a machine-friendly format?

 Encode concepts with (at least) scope note and preferred label

Are there unique local and global identifiers?

Decide on local ID and URI structure

Map encoded file to SKOS

Produce local IDs and URIs and add to encoded file

Share as downloadable files and/or at a SPARQL endpoint.

Build website to serve vocab in a human friendly way

Identifiers are strings of characters which point to a concept (or label). Encoded data often have local identifiers (unique in a local context), but these may not be suitable for Linked Data which require unique identifiers in a global context (also known as URIs). E.g. the URI for the concept of paper in LOD is: http://w3id.org/lof/concept/1481

Mapping means matching the fields of the encoded CSV or XML file to SKOS RDF (as triples). Some of the tools mentioned in note 6 are able to import and convert to SKOS. Manual mapping can be done using tools such as JML, Kemisco and STILETO.

For a large vocabulary many hundreds or thousands of URIs will need to be published and maintained long term (e.g. in LOD the URI for the concept paper is this: http://w3id.org/lof/concept/1481 currently pointing to https://www.ligatus.org.uk/lof/concept/1481). A system to manage these is necessary. URIs may be needed for:
- concepts
- labels
- datasets
- vocabularies

URIs may need to reflect versions. Consider keeping the same URI for a concept in different versions (the essence of a concept should not change), but change the dataset URI based on the version...

The w3id consortium offers a service for URI redirection.
Guidelines for design of URIs can be found online, e.g. here, here and here.

This step is optional: it is useful for humans to be able to look up terms in the vocabulary in addition to sharing as data. Tools to build such a website include BioPortal, Shiremo, Workbench and TemaTema. LCD is discussing the adoption of a repository for SKOSified vocabularies see (?)
Permanent Identifiers for the Web

Secure, permanent URLs for your Web application that will stand the test of time.

The purpose of this website is to provide a secure, permanent URL re-direction service for Web applications. This service is run by the W3C Permanent Identifier Community Group.

Web applications that deal with Linked Data often need to specify and use URLs that are very stable. They utilize services such as this one to ensure that applications using their URLs will always be re-directed to a working website. This website operates like a switchboard, connecting requests for information with the true location of the information on the Web. The switchboard can be reconfigured to point to a new location if the old location stops working.

There are a growing group of organizations that have pledged responsibility to ensure the operation of this website. These organizations are: Digital Bazaar, 3 Round Stones, OpenLink Software, Applied Testing and Technology, Openspring, and Bosatsu Consulting. They are responsible for all administrative tasks associated with operating the service. The social contract between these organizations gives each of them full access to all information required to maintain and operate the website. The agreement is setup such that a number of these companies could fail, lose interest, or become unavailable for long periods of time without negatively affecting the operation of the site.

This website operates in HTTPS-only mode to ensure end-to-end security. This means that it may be used for Linked Data applications that require high levels of security such as those found in the financial, medical, and public infrastructure sectors.

All identifiers associated with this website are intended to be around for as long as the Web is around. This means decades, if not centuries. If the final destination for popular identifiers used by this service fail in such a way as to be a major inconvenience or danger to the Web, the community will mirror the information for the popular identifier and setup a working redirect to restore service to the rest of the Web.

If you would like to add or update a permanent identifier to the website, the preferred procedure is to perform the following steps:

1. Fork the website on Github.
2. Add or update a new redirect entry and commit your changes.
3. Submit a pull request for your changes.

Please help out the maintainers of the service with the following:

- Add contact info in a README or .htaccess comment.
<table>
<thead>
<tr>
<th>Branch</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><code>master</code></td>
<td>Merge pull request #1824 from huanyu-li/master... 30e36c8 16 hours ago 4,470 commits</td>
</tr>
<tr>
<td><code>.assets</code></td>
<td>move to bootstrap 3 and minor cleanup 5 years ago</td>
</tr>
<tr>
<td><code>.utils</code></td>
<td>Added README explaining how the git web hook PHP file works. 7 years ago</td>
</tr>
<tr>
<td>0x1</td>
<td>Update to Juchunko 2 years ago</td>
</tr>
<tr>
<td>360-sfs</td>
<td>Update .htaccess 3 years ago</td>
</tr>
<tr>
<td>3rs</td>
<td>Added link to US EPA Linked Data Services blog post. 3 years ago</td>
</tr>
<tr>
<td>AECLUSsmartApi</td>
<td>Changed parent folder name 16 months ago</td>
</tr>
<tr>
<td>ARK</td>
<td>adding general rules for the ARK project 2 months ago</td>
</tr>
<tr>
<td>AcademIS</td>
<td>AcademIS ontology 2 years ago</td>
</tr>
<tr>
<td>AmlFAIR</td>
<td>repo on GitHub was renamed. Updating the redirect 15 months ago</td>
</tr>
<tr>
<td>BCI-ontology</td>
<td>BCI-O: add turtle resource 3 years ago</td>
</tr>
<tr>
<td>BESDUI</td>
<td>Added entry for BESDUI 4 years ago</td>
</tr>
<tr>
<td>CBGP_SEEK</td>
<td>reconfigured to accept the w3id namespace on the URL 6 months ago</td>
</tr>
<tr>
<td>CONSIDER</td>
<td>Create .htaccess 7 months ago</td>
</tr>
<tr>
<td>DEMETER_ontologies</td>
<td>Update README.md 4 months ago</td>
</tr>
<tr>
<td>DockerPedia</td>
<td>added DockerPedia README file 2 years ago</td>
</tr>
<tr>
<td>EVENTSKG-Dataset</td>
<td>Rename EVENTSKG-Dataset/README.md to EVENTSKG-Dataset... 15 months ago</td>
</tr>
</tbody>
</table>

**About**

Website source code for w3id.org

**Releases**

No releases published

**Contributors**

377

366 contributors

**Languages**

- HTML 66.8%
- PHP 13.2%
How to publish as SKOS
For vocabulary maintainers

START

Is vocab published in a machine-friendly format?

- No: Encode concepts with (at least) scope note and preferred label
- Yes: Are there unique local and global identifiers?
  - Yes: Decide on local ID and URI structure
  - No: Map encoded file to SKOS

Map encoded file to SKOS

Are there unique local and global identifiers?

- No: Produce local IDs and URIs and add to encoded file
- Yes: Share as downloadable files and/or at a SPARQL endpoint

Share as downloadable files and/or at a SPARQL endpoint

Build website to serve vocab in a human friendly way

Encoding is when the concepts and labels from a vocabulary published as text (e.g. in a pdf format) are isolated manually and placed in some sort of data structure, e.g. a comma separated value (CSV) file (typically in a spreadsheet) or an XML file using markup.

Identifiers are strings of characters which point to a concept (or label). Encoded data often have local identifiers (unique in a local context), but these may not be suitable for Linked Data which require unique identifiers in a global context (also known as URIs). E.g. the URI for the concept of paper in LOf is: http://w3id.org/lof/concept/1481

Mapping means matching the fields of the encoded CSV or XML file to SKOS RDF (as triples). Some of the tools mentioned in note 6 are able to import and convert to SKOS. Manual mapping can be done using tools such as JM, Karma and STILETO.

For a large vocabulary many hundreds or thousands of URIs will need to be published and maintained long term (e.g. in LOf) and the URI for the concept paper is this: http://w3id.org/lof/concept/1481 currently pointing to https://www.ligatus.org.uk/lof/concept/1481. A system to manage these is necessary. URIs may be needed for:
- concepts
- labels
- datasets
- vocabularies

URIs may need to reflect versions. Consider keeping the same URI for a concept in different versions (the essence of a concept should not change), but change the dataset URI based on the version.

The w3id consortium offers a service for URI redirection.

Guidelines for design of URIs can be found online, e.g. here, here and here.

Sharing means to make the resulting SKOS file available online (e.g. in a repository as RDF, turtle or RDF/XML) and/or to make the data available at a SPARQL endpoint for querying. LCD is discussing the adoption of a repository for SKOSfied vocabularies see (?).
Linked Conservation Data

https://www.ligatus.org.uk/lcd

Conservation vocabularies repository

This repository hosts raw data representing conservation vocabularies: both published thesauri as well as local wordlists and vocabularies used in institutions or studios. The repository also hosts alignment data between different vocabularies. Best practices for the use of the repository are still being agreed. Current discussions are captured in the Terminology Guidelines document.
LCD repository

• Raw data, one file per vocabulary release

• Short term: engage community to enable conservators to publish

• Long term: fork and pull requests require minimum time from maintainers
How to publish as SKOS
For vocabulary maintainers

1. Encoding is when the concepts and labels from a vocabulary published as text (e.g. in a pdf format) are isolated manually and placed in some sort of data structure, e.g. a comma separated value (CSV) file (typically in a spreadsheet) or an XML file using markup.

2. Identifiers are strings of characters which point to a concept (or label). Encoded data often have local identifiers (unique in a local context), but these may not be suitable for Linked Data which require unique identifiers in a global context (also known as URIs). E.g. the URI for the concept of paper in LOD is: http://w3id.org/lob/concept/1481

3. For a large vocabulary many hundreds or thousands of URIs will need to be published and maintained long term (e.g. in LOD) the URI for the concept paper is this: http://w3id.org/lob/concept/1481 currently pointing to https://www.ligatus.org.uk/lob/concept/1481. A system to manage these is necessary. URIs may be needed for:
   - concepts
   - labels
   - datasets
   - vocabularies

   URIs may need to reflect versions. Consider keeping the same URI for a concept in different versions (the essence of a concept should not change), but change the dataset URI based on the version...

   The w3id consortium offers a service for URI redirection.

   Guidelines for design of URIs can be found online, e.g. here, here and here.

4. Mapping means matching the fields of the encoded CSV or XML file to SKOS RDF (as triples). Some of the tools mentioned in note 6 are able to import and convert to SKOS. Manual mapping can be done using tools such as JM, Kermo and STILETO.

5. Sharing means to make the resulting SKOS file available online (e.g. in a repository as RDF, turtle or RDF/XML) and/or to make the data available at a SPARQL endpoint for querying. LCD is discussing the adoption of a repository for SKOSified vocabularies see (?)..

6. This step is optional: it is useful for humans to be able to look up terms in the vocabulary in addition to sharing as data. Tools to build such a website include: BioPortal, Shsemos, Woobench and TeraTree. LCD is discussing the development of a conservation portal for SKOSified vocabularies see (?)..

START

Is vocab published in a machine-friendly format?

yes

no

Encode concepts with (at least) scope note and preferred label

Are there unique local and global identifiers?

yes

no

Decide on local ID and URI structure

no

how

Map encoded file to SKOS

Share as downloadable files and/or at a SPARQL endpoint.

Build website to serve vocab in a human friendly way

Map encoded file to SKOS

Decide on local ID and URI structure

Produce local IDs and URIs and add to encoded file

how

how
LCD terminology portal

• Terminology querying
• Short term: working with the ResearchSpace project to provide interface
• Long term: ?
Alignment

• Encoding vocabularies
  – to make them machine readable

• Matching terms
  – strategies and methods for alignment
How to match concepts between vocabularies

For vocabulary maintainers

1. Vocabularies are hierarchical when their concepts are linked with broader/narrower relationships (e.g., paper has broader term fibre materials).
   Vocabularies are left hierarchies when narrower concepts are of the same type as their parent concepts. E.g., of isA, in LoB constructed boards are types of boards. E.g. of non-isA, in RSMS the Malachite are bindings are types of bindings, but they are narrower terms of Wooden boards (which are not bindings).

2. Direct mapping: Matching to many vocabularies directly means more matching work, but possibly better coverage (e.g., LoB to RBMS and LoB to Roberts and Emmerling and LoB to etc.).
   Hub mapping: Matching to hub vocabularies like AAT means less matching work but possibly limited coverage (e.g., some LoB concepts do not exist in AAT).

3. Leaf matching is when the top of a hierarchy of concepts from the source vocabulary can be matched to a concept in the target vocabulary which does not contain the same detail.
   Equivalence relationships here can be established when a concept of one vocabulary is marked as exact match of a concept in another, e.g., components is exact match to BBT structural parts of material things.
   Hierarchical relationships can be established when a concept of one vocabulary is marked as broader or narrower of a concept in another, e.g., joint is narrower match to BBT physical features.
   Software tools making this process easier include BBTTab, concept VAT. In general vocabularies need to be SKOSified for these tools to work (see here).

4. Instructions for submitting new concepts to AAT can be found [here].

5. Choosing a vocabulary other than AAT means that it would be particularly relevant to your vocabulary, that it is of about the same detail and that it is likely that people will want to search across the two, e.g., LoB and RBMS.

6. Equivalence relationships can be established when a concept of one vocabulary is marked as exact match or close match to a concept in another, e.g.
   Associative relationships can be established when a concept of one vocabulary is marked as related to a concept in another, e.g.
   Software tools making this process easier include: BBTTab, concept VAT. In general vocabularies need to be SKOSified for these tools to work (see here).

START

Is your vocab strict isA?

Choose BBT

Leaf matching to BBT using hierarchical or equivalence relationships

Decide strategy for vocabs matching

Choose AAT

Submit missing concepts to AAT

Match every concept using equivalence or associative relationships

Choose target vocab for direct mapping

Match through a hub?

Choose target vocab for direct mapping

Store the IDs of the matched target concepts with the IDs of your source vocab

How

How

How

How

How

How

How
How to match concepts between vocabularies
For vocabulary maintainers

1. Vocabularies are hierarchical when their concepts are linked with broader/narrower relationships (e.g., paper has broader term fibre materials).
Vocabularies are also hierarchies when narrower concepts are the same type as their parent concepts. E.g., of isA; in LoB constructed boards are types of boards. E.g. of non-isA; in RSMS the Naughtime were bindings are types of bindings, but they are narrower terms of Wooden boards (which are not bindings).

2. Direct mapping: Matching to many vocabularies directly means more matching work, but possibly better coverage (e.g., LoB to RBMS and LoB to Roberts and Emmering and LoB to etc.).
Hub mapping: Matching to hub vocabularies like AAT means less matching work but possibly limited coverage (e.g., some LoB concepts do not exist in AAT).

3. Leaf matching is when the top of a hierarchy of concepts from the source vocabulary can be matched to a concept in the target vocabulary which does not contain the same detail.
Equivalence relationships here can be established when a concept of one vocabulary is marked as exact match of a concept in another, e.g., components is exact match to BBT structural parts of material things.
Hierarchical relationships can be established when a concept of one vocabulary is marked as broader or narrower of a concept in another, e.g., parts is narrower match to BBT physical features.
Software tools making this process easier include BBTat, concept VMT. In general vocabularies need to be SKOSified for these tools to work (see here).

4. Instructions for submitting new concepts to AAT can be found here.

5. Choosing a vocabulary other than AAT means that it would be particularly relevant to your vocabulary, that it is of about the same detail and that it is likely that people will want to search across the two, e.g., LoB and RBMS.

6. Equivalence relationships can be established when a concept of one vocabulary is marked as exact match or close match to a concept in another, e.g.
Associative relationships can be established when a concept of one vocabulary is marked as related to a concept in another, e.g.
Software tools making this process easier include: BBTat, concept VMT. In general vocabularies need to be SKOSified for these tools to work (see here).

7. Matching IDs of concepts across vocabularies means that the matches remain valid even if the labels for these concepts change.
How to match concepts between vocabularies

For vocabulary maintainers

1. Vocabularies are hierarchical when their concepts are linked with broader/narrower relationships (e.g., parent has broader term than material). Vocabularies are flat hierarchies when narrower concepts are of the same type as their parent concepts. E.g., of non-parent in LoB: constructed boards are types of boards. E.g., of non-parent in RIMB: the Hardware Bindings are types of bindings, but they are narrower terms of Wooden boards (which are not bindings).

2. Direct mapping: Matching to many vocabularies directly means more matching work, but possibly better coverage (e.g., LoB to RBMS and LoB to Roberts and Hennington and LoB to etc.). Hub mapping: Matching to hub vocabularies like AAT means less matching work but possibly limited coverage (e.g., some LoB concepts do not exist in AAT).

3. Leaf matching is when the top of a hierarchy of concepts from the source vocabulary can be matched to a concept in the target vocabulary which does not contain the same detail.

4. Equivalence relationships here can be established when a concept of one vocabulary is marked as exact match or close match to a concept in another, e.g., components is exact match to BBT structural parts of material things.

5. Hierarchical relationships can be established when a concept of one vocabulary is marked as broader or narrower of a concept in another, e.g., parts is narrower match to BBT physical features.

6. Software tools making this process easier include BBTMap, concept: AAT. In general vocabularies need to be SKOSified for these tools to work (see here).

7. Instructions for submitting new concepts to AAT can be found here. Equivalence relationships can be established when a concept of one vocabulary is marked as related to a concept in another, e.g.

8. Associative relationships can be established when a concept of one vocabulary is marked as related to a concept in another, e.g.

Software tools making this process easier include: BBTMap, concept: AAT. In general vocabularies need to be SKOSified for these tools to work.
How to match concepts between vocabularies
For vocabulary maintainers

1. Vocabularies are hierarchical when their concepts are linked with broader/narrower relationships (e.g., paper has broader term sheet materials).
   Vocabularies are flat hierarchies when narrower concepts are of the same type as their parent concepts. E.g., of isA, in LoB constructed boards are types of boards. E.g. of non-isA, in RBSMS the Hardware bindings are types of bindings, but they are narrower terms of Wooden boards (which are not bindings).

2. Direct mapping: Matching to many vocabularies directly means more matching work, but possibly better coverage (e.g., LoB to RBMS and LoB to Roberts and Emrickton and LoB to etc.). Hub mapping: Matching to hub vocabularies like AAT means less matching work but possibly limited coverage (e.g., some LoB concepts do not exist in AAT).

3. Leaf matching is when the top of a hierarchy of concepts from the source vocabulary can be matched to a concept in the target vocabulary which does not contain the same detail.
   Equivalence relationships here can be established when a concept of one vocabulary is marked as exact match of a concept in another, e.g., components is exact match to BBT structural parts of material things.
   Hierarchical relationships can be established when a concept of one vocabulary is marked as broader or narrower of a concept in another, e.g., pants is narrower match to BBT physical features.
   Software tools making this process easier include BBTMap, conceptAAT. In general vocabularies need to be SKOSified for these tools to work (see here).

4. Instructions for submitting new concepts to AAT can be found here.

5. Choosing a vocabulary other than AAT means that it would be particularly relevant to your vocabulary, that it is of about the same detail and that it is likely that people will want to search across the two, e.g. LoB and RBMS.

6. Equivalence relationships can be established when a concept of one vocabulary is marked as exact match or close match to a concept in another, e.g.
   Associative relationships can be established when a concept of one vocabulary is marked as related to a concept in another, e.g.
   Software tools making this process easier include: BBTMap, conceptAAT. In general vocabularies need to be SKOSified for these tools to work (see here).

7. Matching IDs of concepts across vocabularies means that the matches remain valid even if the labels for these concepts change.

START

Is your vocab strict isA?

Yes

Choose BBT

Leaf matching to BBT using hierarchical or equivalence relationships

No

Decide strategy for vocabs matching

Choose target vocab for direct mapping

Match through a hub?

Yes

Choose AAT

Submit missing concepts to AAT

No

Match every concept using equivalence or associative relationships

Store the IDs of the matched target concepts with the IDs of your source vocab
LCD repository

• Alignment data
  – different dataset to actual vocabularies to allow easier versioning
Linked Conservation Data
LCD Terminology Working Group
Vocabulary guidelines

Edited by: Athanasios Velios, Kristen St. John
Contributions by: Anastasia Axaridou, Ceri Binding, Nicola Carboni, Kirsten Dunne, John Graybeal, Ryan Lieu, Joseph Padfield, Eleni Tsouloucha, Jon Ward, Marcia Zeng and others.

Linked Conservation Data is funded by:

Please send comments:
https://www.ligatus.org.uk/lcd/output/193
Future work

• Planning next phase:
  – process vocabularies
    • those which are easy to work with
    • those which are of interest to consortium
    • those ready to adopt LOD
  – populate LCD terminology portal
Thank you

• LCD terminology working group
  – Alberto Campagnolo, Anastasia Axaridou, Ceri Binding, Claudia Marinica, Douglas Tudhope, Eleni Tsoulouha, John Graybeal, Jon Ward, Joseph Padfield, Karen Waldemar, Kristen St.John, Layna White, Marcia Zeng, Maria Theodoridou, Michelle Barger, Nicola Carboni, Ryan Lieu, Stephen Stead